从朗道势垒到奥数 -- 我对未来教育和科学的看法

作者 顾立帆

—— 水往低处流。水可能还会往前流一点，但失去了部分势能，还有各种耗散。

1949 年诺贝尔生理医学奖 leukotomy 存在严重的实验检测错误，并造成灾难性误导。但至今无法纠正。在科学研究上，有错不纠，蔓延开来，会损害科学和文明的基础。

Turing Test 和主流教科书 AI: A Modern Approach 有哲学基础和测试理论上的问题 [1]，该教科书的 Wind Tunnel approach 已造成严重后果；权威学术杂志 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称；破坏了学术语义和科学实验的严谨；Technological Singularity、In Math We Trust、和"人类正在扮演上帝的角色"的观点 [2]，甚至出现在专业研究人士中间；而把脑研究比作登月项目的官方观点，更是惊人的无知 [3]。

这些说明：现实中真正奉行的研究哲学有严重隐患。科学基础和方法上的严重误解，已蔓延到许多方面。科学起源于哲学。科学的基础，不是数学，也不是经验，而是证伪实验。科学实验也不同于经验测试。过分注重技巧，排斥基础理论，排斥对科学的哲学基础的研究分析，会造成科学哲学上的无知和学术上的诡辩盛行。

学术语义和科学实验的严谨一旦被打破，教育、学术研究、和司法等等许多方面，都会出现全局性严重后果，瓦解科学和文明的基础，危及未来的发展 [4]。

科学需要平等讨论和公正实验，来发现和纠正错误，建立新的理论基础，包括自动驾驶的定义和测试问题 [5]。本文从 leukotomy 开始，分析教育和科学研究面临的危机和挑战。

一、Leukotomy 和 BRIAN Initiative

Leukotomy 造成严重问题，原因包括：

1）选择手术对象时没有区分 personalities 和 mental diseases。由于对 personalities 缺乏研究，目前仍无法准确区分两者。这也使得使用基因技术的 BRAIN Initiative 项目更加危险 [6]。

2）给 leukotomy 授与诺贝尔奖时，没有正确地检验 leukotomy 的后果，对提出警告者也没有重视。而本人对 BRAIN Initiative 提出的警告和对 AlphaGo 系列提出的检验要求，也没有被重视。

3）手术对 personality 和 intelligence 有严重损伤。连大规模实施者 Walter Freeman 都在一定程度上承认。由于缺乏对 personality 和 intelligence 的理解，事后的统计无法反映其损伤 [7]。

而无法理解 personality 和 intelligence，是由于缺乏正确描述 personality 和 intelligence 的语言、缺乏开拓科学新学科所需要的哲学性思考和分析，而造成的无知。

另外，leukotomy 获得诺贝尔奖之前的研究，没有作严格控制条件的科学实验，仅仅凭经验性的手术尝试作出主观的评估。这种主观评估跟后来发现的灾难性后果明显不符。

而缺乏严格控制实验条件的主观评估，也正是 Turing Test 的问题之一。Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，也是缺乏严格控制条件的实验情况下的主观评估。因为 superhuman 是针对 generic 的 human，只要有人能设计出证伪的实验，就可以否定该宣称。

Leukotomy 的诺贝尔奖有错不纠，和 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称拒绝严格控制条件的实验证伪，显示 peer review 有严重问题。Peer review 也是主观评估 [8]。

从 leukotomy 开始的问题已经危及脑和智力研究及相关领域，包括 BRAIN Initiative 项目。

二、**Technological Singularity** 和 **AlphaGo** 系列的问题

Nature 杂志上的 AlphaGo Zero 论文，宣称 AlphaGo Zero 取得了 superhuman 的 performance。但 superhuman 的概念跟 generic 的 human 有关。该论文并没有提供任何 superhuman 的证据。

AlphaGo Zero 击败 AlphaGo Master 并不是 superhuman 的证据；电脑围棋胜过一些对计算机和人工智能重要的本质问题尚缺乏了解的棋手，也不是 generic 的 superhuman 的证据。这种 game 只是经验测试，不是科学实验。科学实验需要根据科学原理的分析，设计严格控制条件的针对性实验。

只要有人设计出能证明 AlphaGo Zero 在围棋上不具 superhuman 特质的证伪实验，就可以否定该宣称。科学结论必须接受别人严格控制条件的科学实验的证伪。但 AlphaGo Zero 至今拒绝本人设计的证伪实验。

所以，AlphaGo Zero 的 Nature 论文中的 superhuman 宣称，根本不是科学结论。这个宣称，破坏了学术语义和科学实验的严谨。

科学的基础是证伪，不是经验。逻辑实证主义在科学哲学上是有问题的。因为逻辑判断使用的语言和经验证据之间可能有很大的差别，也就是说经验证据往往不能真正支持逻辑判断。而科学的原理分析，可以发现两者之间的差别。

比如，不管地球上有多少人看到太阳东升西落，也不能证明太阳回绕地球转。不管电脑围棋多少次在 game 中战胜人类，也不能证明电脑围棋有 superhuman 特质。

有再多的经验证据，也不能得出科学结论。还要看别人能不能设计出证伪实验。只要有人能设计出严格控制实验条件的证伪实验，就可以否定经验结论。

近年有人根据人工智能的一些成功经验，炒作 Technological Singularity，实际上是对科学哲学的无知。会误导了脑和智力的研究。

我无意贬低电脑围棋的进展。事实上，我是鼓励电脑围棋的研究。早在 Deepmind 成立之前，我就分析认为电脑围棋可以击败人类顶级棋手，因为围棋是在很小空间里的高度规则化游戏 [9]。这已被证实。但这并不意味电脑围棋有通用智力潜力或超人能力。

AlphaGo Master 击败柯洁后，很多人开始膜拜 AlphaGo Master。而我指出：AlphaGo Master 虽然赢棋，但只要不是最优解，就可能走的是无理棋。这也已被证实，因为 AlphaGo Zero 击败 AlphaGo Master 意味着 AlphaGo Master 很多离谱的走法其实是无理棋。

见到的 AlphaGo Zero 棋谱中，对 AlphaGo Master 的二十局是 AlphaGo Zero 最后版（或稳定版）[10]，有很不寻常现象：AlphaGo Master 执黑的十盘棋中，布局至 16 手完全一样。而 AlphaGo Zero 执黑的十盘棋中，布局至 18 手完全一样，AlphaGo Zero 输棋的 14、16 局，至 45 手完全一样。所以有如下问题：

1）在 Nature Magazime 上发表两篇 AlphaGo 系列文章后，Deepmind 有没有在这二十局棋里把 AlphaGo Zero 和 AlphaGo Master 对局的情况，全面准确地告诉公众？如果答案是，则：

2) 公布的这二十局棋谱中，所出现的布局高度类同，是由于 AlphaGo Zero 接近围棋最优解，还是陷入 local trap？如果认为是前者，需要用严格控制条件的严谨实验来验证。如果是后者，AlphaGo Zero 走的也是无理棋，而 superhuman 的宣称可能对围棋和智力科学研究带来严重误导。

其实 AlphaGo Zero 击败 AlphaGo Master，本身就是人类击败 AlphaGo Master 的证据：人类靠设计新算法系统击败了 AlphaGo Master，而 AlphaGo Master 自身的进化速度远不及人类根据系统结构分析来设计新算法的速度 [10]。Nature 杂志上的 AlphaGo Zero 论文发表前几个月，我就针对 AlphaGo Master 怀疑：利用自我对弈收集数据和 deep-learning 来涨棋，几年后其效果会不会暴露出局限？事实上，AlphaGo Master 的弱点不到四个月就暴露。

即使人类能设计更新的算法击败 AlphaGo Zero [12]，仍不能证明更新的算法有 superhuman 的特质。我设计的证伪实验也可以证伪 AlphaGo Zero 的后继方案。

我验证 AlphaGo 系列有没有 generic human 的围棋智力的实验并不复杂。围棋下得不好的我也可以做实验，只要 Deepmind 同意在实验上合作。目的是找出人工智能技术普遍而严重的问题。我的实验的更大价值在于较准确地测试 AlphaGo 系列的围棋智力上限，这需要一流人类棋手和 AlphaGo 系列对弈，要有使其暴露弱点的方法。而在普通而公平的围棋比赛中，人类一流棋手是否仍有可能击败 AlphaGo 系列，也需要进一步验证。

我设计的实验，实验结果和结果分析会全部公开。别人可以对实验过程和结果作分析、推敲。这才是真正的科学实验方法。如果人类棋手在我设计的公正实验条件下也能胜过 AlphaGo 系列，意味着 AlphaGo 系列所依赖的人工智能技术有严重的缺陷 [13]。这对脑和智力的研究、以及人工智能的发展都有很大价值。

我设计的实验，基于 Turing Machine 和人工智能的局限性，不需要 AlphaGo 系列论文和棋谱中的信息，所以可以验证人工智能和计算机本身的局限性。但出于对科学和公众负责，论文作者应该公布跟论文有关的全面准确的实验结果：包括 AlphaGo Zero 对 AlphaGo Master，和最后版（或稳定版）的 AlphaGo Zero 对 AlphaGo Zero 的更多棋谱，等等。

基于数据的人工智能或统计方法，无法解析未来的模式演变。如果模式演变不稳定，会产生严重误导。有些以前几乎不出现的问题，模式演变后，甚至可能成为大概率事件。要了解智能的模式演变是否稳定，需要研究智力的基础理论 [14]。

三、数学危机和智力的科学研究

科学中虽然大量使用数学，但哥德尔定理实际上已经说明数学不能用来判断科学上的正确性。Turing Machine 及其各种变种，都是数学的子集，当然也不能判断科学上的正确性。不了解 Turing Machine 上的各种问题，夸耀 Universal approximation 只能是误导。

科学起源于哲学。亚里斯多德的物理研究缺乏哲学上的完整性（intergrity）。伽利略澄清了亚里斯多德对力和速度的误解，把物理学奠定在实验基础上，物理研究才变成科学。亚里斯多德的博物学只是经验性的。物种其实是动态发展的。达尔文的演化理论揭示了生命本质规律，上升为哲学理论，才能发展成为科学——生物学 [15]。

达尔文的理论，可以分成两部分：1）演化规律，这是科学理论，可以用实验来验证；2）物种的起源，这是从他的演化规律推导而来的历史理论。历史理论不是严格意义上的科学，历史学上的很多证据远不如科学严谨。宇宙大爆炸理论是从物理学推导而来的历史理论。所以，宇宙大爆炸理论也不是严格意义上的科学 [16]。

科学的基础是 Socratic method 和科学实验，都需要大量使用自然语言来描述、分析和讨论。科学论文也大量使用自然语言和哲学性思考和分析。忽视、排斥科学哲学的人，很可能用一种错误的哲学来看待科学，比如经验哲学或实用主义。

第一次数学危机，其实跟哲学有关。科学哲学在 Karl Popper 以后，基本停滞不前。所以需要新的语言来研究智力现象及其和数学、物理学、生物学的不同和关系。对毕达哥拉斯体系和无理数的引入，以及各次数学危机的本质，也需要重新认识 [17]。

对第一次数学危机引起的智力和知识结构的演变，作重新认识，超越了数学和物理学，涉及科学的本质，是哲学和智力问题。生命需要在热力学的熵定律之上发展，这涉及生命的本质。智力也不同于能量。所以，智力科学也不同于物理学。

Neurosciences 在视觉、听觉、情绪、运动等动物和生理层次的研究上，有很多进展。但对 human specific intelligence 的研究却无法深入，也无法设计出对人工智能的证伪实验。所以智力科学也不同于普通的生物学。

人工智能的研究，从 Alan Turing，到 Geoffrey Hinton，到 Demis Hassabis，技术还在发展，但思考深度和境界却在递减。AlphaGo 系列在算法设计上有进步，在是否超越人类智力这一方向性问题的分析上，却误导了人工智能的发展。

Demis Hassabis 混淆了人类直觉和 deep-learning、reinforcement 等等人工智能技术的区别 [18]，还在 Nature 论文中宣称 AlphaGo Zero 在围棋上有 generic 的 superhuman 特质，破坏了学术语义和科学实验的严谨。作出科学结论前，应该先接受别人的证伪实验。他只主导了一些应用开发，而 Geoffrey 提出了较系统的思考和方法。

Turing Machine 为计算机提供了抽象的理论模型，可以用来分析人工智能、电子计算机、和未来量子计算机的局限性。 Geoffrey Hinton 对计算机的理解，没有达到 Turing Machine 的本质，也就无法超越 Alan Turing 的思想境界，难以理解人工智能的局限性。

重要科学学科的开创，均以哲学性思考为先导，慢慢引入数学。智力现象仍然遵守物理学、化学、生物学的规律。要理解人工智能和计算机的局限性，研究人类和人工智能更好的合作模式，仍然需要哲学性思考为源泉：

一、可以先研究从伽利略的 Dialogue Concerning the Two Chief World Systems [19]，到广义相对论的发展过程，去理解科学规律的本质，理解科学实验和经验测试的不同。

二、还可以研究从波义耳的 The Sceptical of Chemist，到元素周期表，到量子物理的标准模型的发展过程，去了解科学的深入和具体。

三、还可以研究从达尔文的 On the Origin of Species，到遗传学，到生物学综合的发展过程，去理解生命的本质。

在此基础上，发展出更准确的智力概念和语言、哲学基础、和数学形式，来理解 human specific intelligence。需要设计有针对性的实验，来收集少数的临界性和其它本质性信息，来形成和发展结构性理论。这跟 big data 技术刚好相反，也无法由 big data 技术来完成。而 BRAIN Initiative 项目对 big data 方法的过分倾斜，使真正的 human specific intelligence 基础理论研究得不到资源而受排斥。

Big data 是数学，不是科学。不澄清这点，智力科学的新思想就得不到形成的机会。把智力研究转化成科学，可以从可验证的 AlphaGo 系列的问题着手，逐步深入。

智力科学，研究的是智力原理，不是智谋智术。我设计的电脑围棋实验，测试的也不是棋艺，而是智力原理。哲学家和基础科学家，本来就是研究原理的，有长远社会价值，但没有短期效益。跟苏格拉底和伽利略一样，都是弱势群体，也难以说服公众。

诡辩家才是以智术取胜的社会优势者，是能利用社会的智者。但如果以诡辩扭曲学术和科学，控制研究资源，则会从内部瓦解科学和文明的基础。

凡是享受现代教育、现代生活的人，都应该对希帕索斯和伽利略感恩。虽然他们最重要的研究工作，并没有给他们自己带来经济回报，但有这样的人为了追求学术真相而牺牲自己的生活甚至生命，科学才能够持续发展到今天。

四、教育和科学研究中的危机

朗道势垒是苏联物理学家朗道招收学生的最低测试，以难度著称。成功通过朗道势垒本身并不是学术研究的成就。

国际数学奥林匹克（奥数）也起源于苏联的中学数学竞赛，但知识层次较低，针对中学的学生。奥数获奖，虽然是荣誉，但也不是学术成就。象中国的微雕艺术，是有趣而令人尊敬的技艺，但不是学术研究的主要目的。

关注点从朗道势垒到奥数，体现了所关注的知识层次的降低（也是学术境界的降低），和参与人数的进一步扩大。

作为爱好者的活动，奥数可以在年轻学生中提升、培养对数学的兴趣。但把奥数竞赛当作教育的主流指标，却有很多弊端。如果因此过分注重技巧的训练，忽视和排斥对哲学和科学方法的教育和发展，更会破坏教育的全面发展，并危及未来的科学研究 [20]。

科学研究建立在语义和科学实验的严谨上。其需要的一些本质思考和实验方法，从奥数和其他类似的高难度、浅层次、纯技巧训练中是得不到的。

科学哲学在 Karl Popper 以后，基本停滞不前，不但导致智力研究在理论上误入歧途，也使科学研究在整体结构上越来越难以突破。很多耗资巨大的实验结束后，既证明不了什么本质，也否定不了什么本质，使新的科学理论无法成熟、诡辩又开始在学术界流行。

好的科学实验，是要找到实验空间上的临界点，根据实验结果就可以建立新的理论，否定其他可能。但在物理学的很多重要方向，现在要设计出这样的好实验越来越难了。

遗传工程是最近几十年才发展起来的，仍充满活力。但生物基因的新研究方法，有很多也面临准确性、稳定性越来越差的情况。

即使生物基因工程能继续繁荣，也未必能弥补电子和计算机产业饱和所带来的衰退。靠生物基因工程一枝独秀，现在的后工业社会可能要被新的农业社会形态所取代。更理想的社会，是能多方面发展的 [21]。

继经典物理学（各种机械相关产业等等）、化学和凝聚态物理（能源、材料、电子产业等等）、生物学（农业和基因工程等等），下一个关注焦点是：智力科学能否形成并有潜力。而有些智力问题，跟算法复杂性无关。

科学新领域的形成和发展，需要新的思想来纠正现在的错误。Turing Test 只是主观、肤浅的经验测试，缺乏对智力原理的科学分析和针对性测试，所以不是科学实验。对未来的模式演变更无法解析，而处于盲目无知的状态。这种错误测试思维，使现在的人工智能，仍停留在经验方法上，误导了人工智能和智力理论的发展。

逻辑实际上是哲学上的一种初级方法，可以发现语言上的一些问题。亚里斯多德建立逻辑学是为了清除诡辩。但二十世纪的新研究发现，逻辑学本身有问题，并不能真正纠正诡辩问题，甚至会造成其它问题。逻辑无法为数学提供基础，当然更不能判断科学上的对错。所以，科学哲学和科学实验才是分辩对错的依据。科学哲学不但可以为科学理论和实验提供一些思路，而且是确保科学实验结果的有效性和正确解释的基础。

人工智能和神经网络，包括 ResNet 及其变种、Generative Adversarial Networks、Capsule networks 等等，在其流程的许多关键点上，都摆脱不了亚里斯多德哲学的主要问题。

主流教科书 AI: A Modern Approach 第三版在哲学基础和测试理论上的错误，是现在人工智能研究普遍面临的问题，也影响到无人汽车的问题。不加限制说明的 driverless Car，就是 SAE level 5，这需要 human specific intelligence 来处理未来模式的演变。只有深入的基础研究才能设计证伪实验。

在哲学上，“形而上学”和“逻辑实证主义”，这两个翻译过来的中文名词，在语义上都很不严谨，并造成严重误导。而这两种哲学思想本身，也有局限性，不可以作为全局性的哲学基础。

“白马非马”的例子，说明判断语义是否正确，需要涉及抽象概念和对前提条件的分析和界定等等。计算机和人工智能，在分析抽象概念和前提条件等等上，恰恰有本质的局限性。所以，即使能做一些自动翻译，也不能真正理解复杂抽象的语义。而教育、学术研究和司法等等许多重要应用，却需要准确分析语义的正确性。所以会造成严重问题。能作一些自动翻译，却不能真正分析理解复杂抽象语义，就是 Chinese room 的问题。

不理解语言、哲学、数学、科学的发展，就无法真正理解人类智力及其潜力。

这些问题，都源自目前教育和研究中的急功近利，及知识层次和学术境界的急剧下降。在智能挑战的时代，对逻辑和诡辩、图灵机和数学公理系统、科学哲学、和 human specific intelligence 的研究却严重不足，大学也缺乏这方面的课程和教育。这些也是产生 Technological Singularity 和“人类正在扮演上帝的角色”等误导的根源，并对未来的研究和发展，造成全局性的危机 [22]。

科学哲学停滞不前的这二、三十年，正好也是美国平民教育急剧衰退的时期。而科学研究一旦被引入歧途，平等讨论和公正实验的科学基础被破坏，平民教育也会面临更大的灾难。平民教育的普遍衰退，则会造成国力的严重下降。

而开创发展智力新科学，重建平等讨论和公正实验的科学基础，可以为探索多样化的平民教育带来新的机会。

五、数学和科学的演变

从毕达哥拉斯到伽利略，是人类文明走上可持续发展的关键。

毕达哥拉斯的抽象化思维，创立了数学。他对算术和数的定义和体系，按现代抽象代数的概念也是成立的，但不适用于几何和其他一些系统。毕达哥拉斯算术系统本身在科学上是错的，但在数学上又是可以成立的、有价值的、可以继续发展的。所以数学不同于科学，也不能简单用数学来验证科学上的对错。

希帕索斯的发现，本来也是对数学和毕达哥拉斯体系的发展。但希帕索斯事件之后，毕达哥拉斯学派本身就没有太多成就了。其他群体则能继续他们的工作，后来居上。

欧几里德几何创建了重要而又可以完备化的数学理论体系，并带动了阿基米德等等一大批泛希腊化时代的学术研究者。

毕达哥拉斯和欧几里德两千年之后，有牛顿和莱布尼茨创建微积分，罗巴切夫斯基等人推翻欧几里德几何的唯一性，建立非欧几何 [23]。

朗道实际上把自己当作第四流的科学家：牛顿是第一流，爱因斯坦是第二流，创建量子物理的才俊们是第三流，和朗道同时代的美国物理学家理查德·費曼也是第三流。也许朗道是自谦，但他肯定无法和爱因斯坦相比。而俄罗斯历史上是出过门捷列夫的。

不知朗道该如何评价伽利略、达尔文。

按照我对学术成就的评价，创建重要学科、和重要学科中的理论体系，是第一流的工作；在重要学科中提出革命性理论体系，是第二流的工作；对主要理论体系的重要延伸或修补、或开创较小的学术分支，可列入第三流的成就。第一、第二流的学术成就非常罕见，第三流的工作已经非常了不起了。这个标准跟朗道的评价高度吻合。

物理学中，伽利略和牛顿的贡献，跟数学中的毕达哥拉斯和欧几里德相仿，属第一流。相对论和量子力学是重要学科的革命性理论体系，跟数学中的微积分和非欧几何相仿。所以，爱因斯坦的贡献是第二流的，而创建量子力学有许多人的贡献，他们的贡献只能算第三流的。量子电动力学只是小分支，所以理查德·費曼的工作也是第三流的。

牛顿和莱布尼茨在数学中的贡献是第二流的。但非欧几何的创建，由多人逐步完成，有许多人的贡献。罗巴切夫斯基和鲍耶对推翻欧几里德几何的唯一性最重要，他们和希帕索斯（发现无理数）、亚里斯多德（逻辑学）、帕斯卡（概率论）、哥德尔（哥德尔定理）相仿，仅略逊于牛顿和莱布尼茨。

拉格日朗（数学分析的成熟化）、图灵（图灵机作为计算模型）的贡献是第三流的；数学中的高斯，和物理学中的朗道，能不能挤进第三流，我不知道；

祖冲之的贡献，未必能超过高斯。但有观点认为《九章算术》首次引入负数 [24]。此说如果确实，《九章算术》的数学成就，也许并不比希帕索斯和亚里斯多德逊色很多。这可能跟中国古代的阴阳玄学哲学观念有关 [25]。

数学是非常古老的学科，现在看起来很重要的一些成就，跟数学历史长河中的很多大师比起来，不一定更重要。十七、十八世纪的数学，和二十世纪前半期的物理学类似，已经十分兴旺，但基础是更早的人奠定的。

达尔文跟毕达哥拉斯和伽利略相仿。孟德尔不如欧几里德和牛顿，因为生物学整体理论在不久之后还需要几次重要的综合才能形成。

波义耳和门捷列夫在化学中的作用，有点类似于物理学中的伽利略和牛顿。但门捷列夫和波义耳不但时间相隔较久，而且中间有很多人的重要贡献。所以，他们的学术贡献，也不如伽利略和牛顿。

这里的评价，全部按照学术贡献的大小。不是评估个人智力，而是评估对人类整体智力发展的影响。人各有特长，个人智力也不一定都表现出来，很多情况下是不可比较的。而学术贡献，是已完成的工作，可以客观分析。

整理历史上学术贡献的大小，一方面是为了研究人类智力的演变和其中重要的临界点，另一方面也是研究能促进未来发展的各种因素。现在把智力研究转化成科

学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。

如果物理学、化学、生物学、人工智能或现有数学等等，都不能提供这些概念、基础、和形式，就需要创建新的学科。能不能提供，可以通过证伪实验来验证。

科学和数学的本质和方法，有所不同。科学兴起虽然较迟，但能弥补数学思维方法的一些严重不足。事实上，物理科学比数学更依赖于哲学性思考。而生命科学的哲学性，比物理科学更重。开创智力新科学，比起生命科学也更具哲学上的挑战性。

从数学、到物理科学、到生命科学、到智力科学，随着哲学性思维比重的不断加大，其研究和教育也会越来越平民化，可以为普通人带来更多的机会。因此也需要探索智力科学时代平民化教育和研究的新方式。

新的哲学思维，也会带来思想上的新挑战。而希帕索斯和伽利略因为追求学术真相而受迫害，更应引起人们的警觉。相比之下，高斯虽是优秀的数学家，却缺乏追求学术真相来推动进步的求道精神，历史地位有限。

公正的研究环境、和免受学术迫害，是科学能够持续发展的前提。

六、结语

把智力研究转化成科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。科学的生命力在于设计可验证的理论。经验测试不是科学实验，再多的经验结果也不能导出科学结论。科学结论必须接受别人严格控制条件的科学实验的证伪。

对智力的基础科学研究，不但可以帮助人类和人工智能更好地配合，而且可以带动脑科学、语言学、心理学、哲学、数学等等一大批学科的发展。也将为多样化的平民教育带来新的机会。

Leukotomy、Turing Test、主流教科书 AI: A Modern Approach 中的 Wind Tunnel approach、Technological Singularity 和“人类正在扮演上帝的角色”的观点，权威学术杂志 Nature 上的 AlphaGo Zero 论文中的 superhuman 宣称，都破坏了学术语义和科学实验的严谨。

学术语义和科学实验的严谨，一旦被破坏，在教育、学术研究、司法等等许多方面，会引发全局性的危机。知错不改，危机会越演越烈，瓦解科学、文明、财富的基础。

科学上的正确与否，不是凭主观去判断该相信谁，而是靠严谨设计、严格控制条件的实验来验证。而 Technological Singularity 和 AlphaGo Zero 的 superhuman 宣称，超出了人工智能开发和围棋竞技的范围，所以要用严格控制条件的科学的新实验来证伪 [26]。

在科学上观点正确却没有权势没有资源的一方，要求公正的实验是最后的生机。如果有权势资源的那方，拒绝以公正的实验来检验，等于要把正确观点逼入绝路。所以，权势和资源也应该意味着责任。

~~~~~~~~~~~~

[1] AI: A Modern Approach 第三版把苏格拉底和柏拉图从哲学基础中删去，而把亚里斯多德哲学当成正确思维的基础：“Aristotle... was the first to formulate a precise set of laws governing the rational part of the mind.”(On page 5)。

Galileo Galilei 在 Dialogue Concerning the Two Chief World Systems 中已经指出了亚里斯多德哲学并不是科学的理性思维。

Immanuel Kant 对人和思维作过深入研究，他的哲学已经远远超出亚里斯多德哲学。所以哲学上的理性思维，实际上也无法用亚里斯多德哲学正确表达。
~~~~~~~~~~~~

Gödel's theorems 更说明了逻辑实证主义有问题，亚里斯多德哲学甚至不能完整表达数学中的理性思维。

所以，亚里斯多德哲学在哲学、数学、科学上都不能准确表达理性思维，当然不可能是 "a precise set of laws governing the rational part of the mind"。

这个问题不澄清，不但人工智能研究受到误导，脑科学、语言学、心理学、哲学等等一大批学科的发展会受到阻碍。越来越多的领域试图用非理性行为来解释现象，实际上很多是出于对理性的误解，看不清非理性表象下面真正的理性原理。

该教科书也没有区分清楚 Turing Test 和科学实验的不同，实际上把经验测试当作人工智能的检验方法，而对人工智能和智力理论的发展有误导。

[2] In God We Trust 是美国的格言。这里提到的 God 是全能神。现代科学可以构造多个具有哲学完整性且无法证伪的科学理论。这些无法证伪的科学理论，使全能神的存在，也无法被证伪。但人不是全能神，人无法完整理解全能神，所以只能 Trust。

Occam's Razor 是一种较陈旧的哲学理论，不但不是科学，也不是严谨的科学哲学。所以从几种无法证伪的理论中，选择最简单的理论，并不是科学结论。对奥卡姆剃刀在学术研究中的广泛使用，需要科学哲学的新研究来澄清其隐含限制、限定其使用。

科学理论的简单和复杂，甚至不一定存在数学上的全序关系。所以，几种无法证伪的科学理论中，也不一定存在最简单的理论。

数学是人的一种智力理论，不是全能神。数学的局限性，早已被证明。所以，In Math We Trust 并不正确，是迷信。In Math We Trust 在精英专业研究人士中的流传，暴露了科学哲学教育的缺乏和现代教育的问题。

Universal approximation theorem 的名称，也很容易误导。定理中所能趋近的函数，是有严格限制条件的。并不是所有函数都能无限趋近。

[3] 登月项目是工程。需要定义明确的实施目标。而脑研究是基础科学研究，需要通过科学实验的证伪来不断修改假设。所以，工程和基础科学研究的方法完全不同。出现这样的混淆，是教育和研究严重失败所造成的无知。

[4] 在政治宣传、商业宣传中，诡辩和语义的混淆比比皆是，难以清除。所以，科学研究和司法，需要跟政治宣传和商业宣传严格区分，保持应有的独立性。

[5] SAE level 4 的自动驾驶，在不同的驾驶区域，实际上有不同的技术要求。所以这个定义，会隐含很多问题。即使在同一区域，未来不稳定的模式演变，也可能使已通过 SAE level 4 测试的自动驾驶系统，出现严重问题。

科学需要平等讨论和公正实验。平民因此可以靠科学上的努力，弥补一些自己的弱势。生命科学的开创人达尔文、孟德尔等人，并不是精英色彩很重的学者。而开创智力科学新领域，探索新的教育和研究方式，可以为多样化的平民教育带来更多的机会。

[6] 科学研究中使用什么技术，应该根据研究需要来选择。BRAIN Initiative 先选择技术方法，由技术方法来决定科学研究的途径，是对科学研究的误导。证伪实验的宗旨，是要收集少数的临界性和其它本质性信息，来形成或改变整体性知识结构。这跟基于 big data 技术的 BRAIN Initiative 的思路刚好相反，当然也无法由 big data 技术来完成。

BRAIN Initiative 用来争取巨额研究经费的不但包括精神分裂症、自闭症这种跟正常人的错乱思维、自闭行为界线模糊的 mental diseases，也包括 anxiety disorders like depression and post-traumatic stress disorder, obesity and eating disorders, bipolar disorder, and mental retardation 等等。那些 disorders 是否是 diseases，争议很大。

所以，BRAIN Initiative 没有区分清楚 personalities 和 mental diseases，却把大量宝贵的政府研究经费耗费在非常盲目的 Big Data 技术上，使更有价值的研究反而得不到资源。

[7] 按政府报告 Leucotomy in England and Wales, 1942-1954，该手术的 9284 名患者中，41%恢复或有明显改善，28%稍微改善，25%显示无变化，2%恶化，4%死亡。

但这个统计很容易被误解。如此破坏性的脑手术，当然会破坏 personality 和 intelligence。所以，这里的 25%显示无变化，不是指 personality 和 intelligence 未受破坏，仅指为控制精神状态而特殊定义的 clinical condition 无变化。同理，41%恢复或有明显改善和 28%稍微改善，也只是特殊定义的 clinical condition 有改善，personality 和 intelligence 实际上都有损伤。

本来，leucotomy 手术就是为了控制精神状态，而不惜损伤脑功能。

也有冠名 Renato M.E. Sabbatini 教授的文章指：Even lobotomy's preponents admitted that only one third of the operated patients would improve, while one-third remained the same, and one-third got worst 。这些数据来源不明，跟 Leucotomy in England and Wales, 1942-1954 的差距也很大。（http://www.cerebromente.org.br/n02/historia/lobotomy.htm）

但这里的 one third would improve 和 one-third remained the same，也仅仅指特殊定义下的 clinical condition，不是指 personality 和 intelligence 未受破坏。

所以，由于缺乏正确描述 personality 和 intelligence 的语言，就无法对 leucotomy 手术和使用基因技术的 BRAIN Initiative 方法的破坏性后果作研究。这种情况急需改变。

[8] 对学术论文的审核，只代表审核者的观点，不是学术上的定论。所以，学术上的苏格拉底方法很重要。要给被审核者和其他人反驳的机会。

对科学论文的审核，也跟数学论文不同。数学论文只能靠 peer review。科学论文中的结论，可以用新的证伪实验来推翻。其错误，从论文本身可能看不出。所以，对科学论文的 peer review，并不能保证其正确性。要让别人有机会设计新实验来证伪。

对科学论文的审核，可以发现推理问题，实验数据是否真实等等。但推理问题只能引起注意，有可能是出于理解有误。所以下结论要经过对话、反驳、和实验证伪。AlphaGo Zero 论文中的 superhuman 宣称，是针对 generic 的 human，并没有相应的证据。本人却可以设计严格控制条件的实验来证伪此宣称。但 AlphaGo Zero 方面至今拒绝对话和实验。

[9] 许峰雄在 Cracking Go 文中更早提出。但他的依据很多来自开发 Deep Blue 的经验，对围棋不一定成立。我的理论分析，不依赖经验，反而能成立。我的理论分析还可以进一步延伸：AlphaGo 系列虽然赢棋，但只要不是最优解，就可能走的是无理棋；可以设计严格控制条件的严谨实验来检验 AlphaGo 系列是否有通用智力潜力，是否超人。

[10] 根据 http://www.alphago-games.com/, AlphaGo Zero 对 AlphaGo Zero 中只有一局是最后版（或稳定版）的。在 https://www.101weiqi.com/chessbook/player/38348/也有相关棋谱。

[11] AlphaGo Master 不是生命，AlphaGo Master 进化快慢，也跟生物的演化本质无关。部分原因是裁定围棋胜负的规则是由人类定义的，而生物演化的规则并非由人类定义。

[12] http://www.alphago-games.com/上，Full Strength of Alphago Zero, i.e. Final Form，被称为 40 Blocks。20 Blocks 被称为 Not Full Strength of Alphago Zero。有更新结构的系统能击败 Alphago Zero 也是非常可能的。

[13] 我选择围棋作实验工具，是因为可以用围棋验证人工智能和计算机的一些基础科学理论，跟军事战略和战争艺术无关。围棋是和平的竞技，高度抽象而又直观。既古老又现代，能吸引亚洲以外的电脑研究。已经成为跨文化研究、交流的平台。

第一轮实验后，开发者可能修改软件。所以有些实验证据只能在第一轮实验中收集。第一轮实验需要跟最强版本电脑围棋作较大规模的实验，以收集足够的实验证据。

我对 AlphaGo 系列的实验不是白盒测试，不涉及知识产权。如果 Google 愿意公正地检验，可以仿照美国国会图书馆先例，先向国会图书馆递交 AlphaGo 柯洁版、AlphaGo Zero、和最新 AlphaGo 版本所需要的所有硬件、软件和文档，能捐款给图书馆用于管理和维护则更好。

以后在第三方场地实验时，可以向国会图书馆核对软件、硬件和文档等版本。Human level artificial intelligence 的检验，是人类发展史上的里程碑。如果无人能对 AlphaGo 系列证伪，国会图书馆就有继续保留其版本的价值。其他人有实验想法，可以先做。如果其他人都失败了，就没有道理拒绝我的实验方案。

这也为美国政府了解人工智能的最新进展提供了渠道，并为以后的人工智能测试建立了一个范式。一旦相关实验设施建立好并熟悉了这种流程，以后重要的人工智能测试费用会大大降低。

[14] 根据新闻报道，2012 年时，就有盲人使用自动驾驶汽车；2015 年时，有盲人单独使用无人汽车。根据公开的事实，已经知道当时无人驾驶汽车没有成熟，对人工智能和无人驾驶汽车有误判和误导。也说明之前的经验测试不可靠。

我以前写过:“Go gaming is strictly defined within a very small space. Industrial automations are typically designed in well controlled environments, but not strictly defined. Car driving is regulated, but the environment is not well controlled”。无人驾驶汽车中的问题，远比电脑围棋复杂。而即使在电脑围棋中，仍有问题可以被证伪。

另外，无人驾驶汽车的测试，完全不同于其他汽车。汽车中的动力（油动或电动）、机械、电子等设计，早就具备坚实的科学基础。而无人驾驶的人工智能，没有科学基础，也未经严格控制条件的科学实验验证，危险性很大。

现在使用的是经验测试，无法解析未来模式的演变。而普通人可能需要很多年跟无人驾驶汽车的大规模互动，才能接触到其临界点，引发重大的模式演变。所以需要对智力的基础理论作前瞻研究，设计新的严格控制条件的科学实验来验证和证伪。

[15] 哲学理论和科学理论，需要哲学上的完整性（intergrity）。其他领域的工作，不一定达到这种标准。像数学上的欧几里德几何和非欧几何，同时存在。但数学可以作为科学的工具，本身并不是科学。事实上，非欧几何的出现，就明确宣告欧几里德几何不是科学，而仅仅是数学。哥德尔定理更说明数学不能判断科学上的正确性。

[16] 历史理论不是严格意义上的科学，并不意味历史理论就一定是错的。只是说历史理论达不到科学理论的标准。

[17] 物理学界有人质疑时空是连续的。有人更以此质疑我的智力理论。对我的质疑毫无道理。第一，要看非连续的时空理论是否可验证。很多现代物理理论根本没法验证，没法验证的就不是科学。象围棋走法的数量超过宇宙的原子数就是这种理论，因为对宇宙原子数量的推测根本不可靠，也无法验证。第二，著名物理学家的观点不一定是对的，如牛顿迷恋炼金术。第三，时空的定义和时空的测量是两回事，我的智力理论跟时空在测量上是否是连续根本没有关系，我研究的是无理数在科学及其未来发展上是否不可缺少，这是智力科学的问题。这再次说明智力科学和物理学是不同的。

[18] 直觉只是人类智力中的一部分。郎道和理查德·费曼就是非常注重直觉的物理学家。下文中提到他们并不是最优秀的物理学家，他们自己也承认。

而 deep-learning、reinforcement 等人工智能技术显然不具有人类直觉。我设计的对 AlphaGo Zero 的实验就可以验证这一点。

[19] Dialogue Concerning the Two Chief Word Systems 实际上把科学建立在 Socratic Method 和科学实验的基础上。文中有错误，特别是关于潮汐的。科学不是完美的，但科学需要能改进错误的方法。文中倡导的思考和方法，可以发现和纠正关于潮汐的错误。

伽利略文中的错误，和第谷的错误不一样。伽利略的思想方法是对的，理论体系也是对的。而第谷的思想方法是错的，理论体系也是错的。如果采用第谷的理论，就无法发展出牛顿的物理基本理论。混淆人工智能和人类智力，也无法发展出智力科学。

第谷的观察工作仍有价值是侥幸，因为地心说、日心说、第谷理论在观察方法上没有争议。而很多科学实验，需要严格控制实验条件。错误的实验条件会导致错误的结论。

所以该文提供了科学的良好范本。科学不应该戴上圣人或权威的面具，科学应该放下身段来平等讨论。

[20] 近二、三百年，俄罗斯的人口规模和 talent pool 的规模远超过其他欧洲国家，在人才的衡量和选拔上有优势。中国的人口规模更大，远超过世界上的其他发达国家。奥数在中国受到追捧，也不难理解。但靠人才优势以技巧争锋，也有弊端。

美国人口规模也不小，近几十年更有财力在全世界范围吸引人才。所以，过去几十年，美国在 talent pool 上的规模优势，其实比中国更强。美国以前在量子物理、天体物理、遗传学、计算机工程等需要新思想的学科上，比前苏联强。但近年则越来越偏向技巧，排斥思想，也忽视和放弃了以往让人有机会做最擅长工作的传统。所以在走下坡路。

[21] Personal computers、Internet、smartphones 的发展是建立在以前重要科学进步的基础上。搭顺风车创业当然比较容易。像 digital cameras、GPS、smart wearables、virtual reality、quantum computer 的市场就炒不大。

无人驾驶汽车的危险性在于大规模应用中对未来模式演变的无知。小规模应用的场景和风险容易控制，但对未来发展也没多少影响。

[22] 根据哥德尔定理，图灵机和数学本身都有严重的不足。Universal approximation 并不存在。Technological Singularity 只是欲望，并无科学依据。AlphaGo Zero 的 superhuman 宣称则可以通过严格控制条件的严谨科学实验来证伪。

[23] 高斯不但没有公布他自己的非欧几何想法，也没有公布 1819 年就交给他的 Ferdinand Schweikart 的类似想法。所以高斯对非欧几何，不但无功，反而有过。原因就在于高斯太注重坚实的技巧性，不了解尚未成熟的思想体系的价值，也不重视苏格拉底方法对学术发展的作用。高斯不但缺乏学术眼光，也达不到罗巴切夫斯基的思想境界和层次。

罗巴切夫斯基和鲍耶推翻欧几里德几何的唯一性，已是 1830 年前后了；他们在非欧几何上贡献最大。黎曼几何在物理相对论研究中，有价值。但黎曼在非欧几何中，只是作延伸性工作，贡献次于罗巴切夫斯基和鲍耶。贝尔特拉米为非欧几何带来新的语义和完整性，也很重要。

由于被高斯埋没，Ferdinand Schweikart 可能没有对非欧几何的发展产生多少实际影响。我也不知道如何评价他的工作。

[24] 《九章算术·方程》：“正负术曰：同名相除，异名相益，正无入负之，负无入正之。其异名相除，同名相益，正无入正之，负无入负之。”

也有观点认为，《九章算术》之前，中国已经出现负数概念。

[25] 中国古代百家争鸣时期，就有阴阳家，《道德经》和解释《易经》的《十翼》也都论及阴阳。阴阳概念的真正起源，有待考证。但《易经》体系的阴阳并不导出负数，而是：老阳数九，老阴数六，等等。

阴阳概念，不但可以很容易导出正负数，也为现代正反物质的理论提供了一个很自然的通俗理解语言和思维模式。但能相互抵消的正负数，是在阴阳概念基础上的飞跃。现代正反量子湮没的理论，又是对可抵消的正负数概念的飞跃，因为量子湮没是伴随有能量形态的转化的。从中正好可以看到从哲学、到数学、到科学的思维模式的演变和发展。

但阴阳不是科学哲学语言，进一步发展的潜力有限。在形成正负数以后，中国古代数学也没有在这个方向有多少后继发展。所以，即使哲学能帮助形成一些新的思想，哲学本身也需要不断发展，需要新的语言概念。“道可道，非常道。名可名，非常名。”，至少说明不能固步自封。拒绝别人的质疑和证伪实验，也是在拒绝进一步发展的机会。

中国古代没有像希帕索斯那样对无理数的准确发现，对类似无理数的概念没多少抵触，但也不理解其价值。这里有哲学和思维模式的不同，受此影响智力的发展轨迹也不同。

[26] 人工智能最大的危险，不是其太强大可能控制或消灭人类，而是大规模应用不成熟的人工智能可能带来巨大灾难。所以，能通过严格控制条件的电脑围棋科学实验来前瞻研究人工智能技术的局限性，研究智力基础理论问题，对人类是一种幸运。